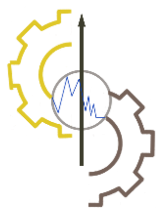
E-ISSN: 2717-4581 | P-ISSN: 2783-1345



Innovation Management and Operational Strategies

www.journal-imos.ir

Inov. Manage. Oper. Stra. Vol. 2, No. 2 (2021) 180–194.





Paper Type: Original Article

A Clustering of Investors' Behavior according to Their Financial, Behavioral, and Demographic Characteristics (An Application of K-means Algorithm)

Marziyeh Nourahmadi[1,] , **Hojjatollah Sadeqi** [2,*]

[1] *Department of Financial Engineering, Faculty of Economic, Management and Accounting, Yazd University, Yazd, Iran; mnourahmadi@ut.ac.com.*
[2] *Department of Finance and Accounting, School of Management and Economics, Yazd University; sadeqi@yazd.ac.ir.*

Citation:



Nourahmadi, M., & Sadeqi, H. (2021). A clustering of investors' behavior according to their financial, behavioral, and demographic characteristics (an application of K-means algorithm). *Innovation management and operational strategies,* 2(2), 180-194.

Received: 21/04/2021 Reviewed: 29/05/2021 Revised: 19/06/2021 Accept: 10/07/2021

Abstract

Purpose: One of the issues that significantly impact how people invest is the behavioural characteristics of investors. Given the importance of this issue, investors should be able to categorize investors into different classes and recommend investments appropriate to the personality type of the same class for each class. One of the solutions that can be used for this purpose is clustering. Clustering is one of the unsupervised learning methods and has a descriptive nature. In this method, the data are allocated based on a similarity criterion so that the data in each cluster are most similar and the least comparable to the data in other clusters.

Methodology: This study identifies a group of investors with similar ability and willingness to accept risk using K-means clustering and Affinity propagation clustering. We also show how to allocate assets effectively using investor characteristics and clustering techniques.

Findings: Use silhouette coefficient to evaluate two clustering methods to select the best method for data clustering. The k-means coefficient was equal to 0.17, and the Affinity propagation clustering was equal to 0.097. Therefore, we choose the k-means method as the optimal clustering method. Using the K-means clustering method, we cluster investors based on financial, behavioural, and demographic characteristics, and according to the clustering results, we divide individuals into seven categories with low to high-risk acceptance.

Originality/Value: All calculations in this study were performed by Python 3.8. Investment managers and stock advisors can use the results of this study.

Keywords: Clustering, Investment consulting, Risk tolerance, Optimization, Investor behavior, Recommender system.



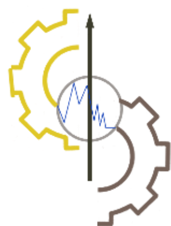
Corresponding Author: sadeqi@yazd.ac.ir



http://dorl.net/dor/20.1001.1.27831345.1400.2.2.6.4



 https://doi.org/10.22105/imos.2021.289811.1111



Innovation Management & Operational Strategies

مدیریت نوآوری و راهبردهای عملیاتی

دوره ۲، شماره ۲، (۱۴۰۰)، ۱۹۴-۱۸۰

www.journal-imos.ir





نوع مقاله: پژوهشی

خوشه‌بندی رفتار سرمایه‌گذاران بر اساس ویژگی‌های مالی، رفتاری و جمعیت شناختی: مطالعه‌ای بر اساس الگوریتم K-means

مرضیه نور احمدی[1,*]، حجت‌الله صادقی[2]

[1]گروه مهندسی مالی، دانشگاه یزد، یزد، ایران.

[2]گروه حسابداری و مدیریت مالی، دانشگاه یزد، یزد، ایران.

چکیده

هدف: یکی از موضوعاتی که تأثیر بسزایی در نحوه سرمایه‌گذاری افراد دارد، خصوصیات رفتاری سرمایه‌گذاران است. با توجه به اهمیت این موضوع، سرمایه‌گذاران باید بتوانند افراد سرمایه‌گذار را به طبقات مختلف دسته‌بندی نمایند و برای هر طبقه، سرمایه‌گذاری متناسب با تیپ شخصیتی همان طبقه را به آن‌ها توصیه کنند. یکی از راهکارهایی که برای این منظور می‌توان استفاده نمود، خوشه‌بندی است. خوشه‌بندی یکی از روش‌های یادگیری بدون نظارت بوده و ماهیتی توصیفی دارد. در این روش داده‌ها بر اساس یک معیار مشابهت به‌گونه‌ای تخصیص می‌یابند که داده‌های موجود در هر خوشه دارای بیشترین شباهت با یکدیگر و کمترین شباهت با داده‌های موجود در خوشه‌های دیگر باشند.

روش‌شناسی پژوهش: در این پژوهش با استفاده از خوشه‌بندی *k-means* و روش انتشار وابستگی، مجموعه‌ای از سرمایه‌گذاران با توانایی و تمایل مشابه برای پذیرش ریسک را شناسایی می‌کنیم. همچنین نشان می‌دهیم که چگونه می‌توان با استفاده از تکنیک‌های خوشه‌بندی به تخصیص مؤثر دارایی‌ها با استفاده از خصوصیات سرمایه‌گذاران پرداخت.

یافته‌ها: از ضریب *silhouette* برای ارزیابی دو روش خوشه‌بندی استفاده نموده تا بهترین روش را برای خوشه‌بندی داده‌ها انتخاب نماییم. ضریب *k-means* برابر با ۰/۱۷ و انتشار وابستگی برابر با ۰/۰۹۷ شد؛ بنابراین ما روش *k-means* را به‌عنوان روش خوشه‌بندی مطلوب انتخاب می‌نماییم. با استفاده از روش خوشه‌بندی *K-means* به خوشه‌بندی سرمایه‌گذاران براساس ویژگی‌های مالی، رفتاری و جمعیت شناختی پرداخته و با توجه به نتایج خوشه‌بندی، افراد را به هفت دسته با پذیرش ریسک کم تا زیاد تقسیم‌بندی می‌نماییم.

اصالت/ارزش‌افزوده علمی: کلیه محاسبات این پژوهش توسط پایتون ۳/۸ انجام‌شده است. نتایج این پژوهش می‌تواند توسط مدیران سرمایه‌گذاری و مشاوران توصیه گر سهام مورداستفاده قرار بگیرد.

کلیدواژه‌ها: خوشه‌بندی، مشاوره سرمایه‌گذاری، تحمل ریسک، بهینه‌سازی، رفتار، سیستم‌های توصیه‌کننده.

* نویسنده مسئول

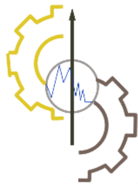
sadeqi@yazd.ac.ir

https://doi.org/10.22105/imos.2021.289811.1111

http://dorl.net/dor/20.1001.1.27831345.1400.2.2.6.4

۱- مقدمه

اطلاعاتی که ما تجزیه‌وتحلیل و استفاده می‌نماییم، به‌سرعت در حال افزایش است. داده‌کاوی فرایند استخراج داده‌های مربوط از حجم زیادی از داده‌ها است، همچنین روش کشف و پیدا کردن الگوی مناسب از حجم زیادی از مجموعه داده‌ها است. هدف اصلی از فرایند داده‌کاوی این است که اطلاعات مناسب و مرتبط از مجموعه داده‌ها استخراج شود و آن را به ساختار قابل‌فهم تبدیل نماید. یکی از زیرمجموعه‌های داده‌کاوی، سیستم توصیه است. سیستم‌های توصیه این قابلیت را دارد که بر اساس تاریخچه و نمایه کاربران پیشنهاد کند که آیا یک کاربر آیتم خاصی را ترجیح می‌دهد یا نه.



سیستم‌های توصیه‌کننده به‌صورت کلی دو نتیجه دارند:

– به کاربر در تصمیم‌گیری کمک می‌نماید (به‌طور مثال، از میان چندین گزینه پیش رو کدام انتخاب بهتر است و...).
– باعث افزایش آگاهی کاربر، درزمینهٔ موردعلاقه وی می‌شود (به‌طور مثال، در حین ارائه پیشنهاد کاربر با اقلام و اشیاء جدیدی که قبلاً آن‌ها نمی‌شناخته، آشنا می‌شود).

یکی از دغدغه‌های مدیران انتخاب سرمایه‌گذاری در دارایی‌هایی است که مطابق با ترجیحات مشتری نسبت به ریسک و بازده است. در روش مشاوره روبو[۱] ، فرایند سنتی سرمایه‌گذاری انتخاب نوع دارایی و وزن دارایی‌ها را با استفاده از الگوریتم‌ها بر اساس ریسک‌پذیری سرمایه‌گذاران به‌طور خودکار و اتوماتیک انجام می‌دهد. برای اطلاعات ورودی الگوریتم‌ها، سرمایه‌گذار به‌صورت آنلاین در مورد تحمل ریسک خود به پرسشنامه پاسخ می‌دهد (استاتمن[۲]، ۲۰۱۸). یکی از ابزارهایی که می‌تواند به روبوادوایزرها در جهت شناسایی تیپ‌های شخصیتی مختلف سرمایه‌گذاران کمک نماید، خوشه‌بندی سرمایه‌گذاران است.

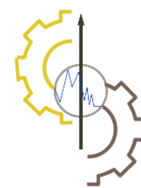
مشاوره روبو از سال ۲۰۰۶ آغازشده است و راهی جدید برای مدیریت پرتفلیو است. روبوادوایزرها پلتفرم‌های دیجیتالی هستند که خدمات برنامه‌ریزی مالی مبتنی بر الگوریتم و خودکار را مانند سرمایه‌گذاری ارائه می‌دهند. این سرویس به خاطر توانایی خود نیاز به نظارت انسانی ندارد. روبوادوایزرها اطلاعاتی را از مشتریان در مورد وضعیت مالی خود و همچنین اهداف آینده خود جمع‌آوری می‌کند. برای انجام این کار، از شما خواسته می‌شود که به یک نظرسنجی یا چند سؤال آنلاین پاسخ دهید. از این داده‌های واردشده برای ارائه مشاوره استفاده می‌کنند. اینکه چه سرمایه‌ای باید انتخاب کنم؟ چه مقدار پول باید در دارایی‌های مختلف سرمایه‌گذاری کنم؟ زمانی که ارزش سرمایه‌گذاری کم شد چه کاری باید انجام داد؟ این موارد سؤالاتی است که اکثر افرادی که تاکنون سرمایه‌گذاری کرده‌اند و یا قصد سرمایه‌گذاری دارند، از خود پرسیده‌اند.

شرکت‌های خدمات مالی از قبیل بانک‌ها، کارگزاری‌ها، ادارات خانوادگی، شرکت‌های بیمه زندگی و تراست‌ها، خدمات سرمایه‌گذاری را به مشتریان خود ارائه می‌دهند و به آن‌ها کمک می‌کنند تا به اهداف خود دست یابند. چنین خدمات سرمایه‌گذاری معمولاً شامل مشاوره در استراتژی‌های سرمایه‌گذاری، مدیریت پرتفلیو اختیاری که در آن مشتریان، مدیریت پرتفلیو را به خبرگان این حوزه از بازارهای مالی محول می‌نمایند. توصیه استراتژی‌های سرمایه‌گذاری مالی یک کار پیچیده است. عمدتاً، مشاوران مالی باید قبل از اینکه گزینه‌ای پیدا کنند که همه نیازها و محدودیت‌های سرمایه‌گذاران را برآورده نماید با مشتریان ثروتمند خود گفت‌وگو کنند و از بین چندین پیشنهاد سرمایه‌گذاری، مناسب‌ترین آن را پیدا نمایند. درواقع، شناختن مشتریان و ارائه پیشنهادها سرمایه‌گذاری شخصی به آن‌ها، امروزه یک جنبه ضروری از یک استراتژی مشورتی سودمند و مؤثر است. سرمایه‌گذاری‌های فناوری اطلاعات در جهت افزایش شفافیت، ارائه گزارش بهتر و به‌موقع به مشتری است اما بر روند تصمیم‌گیری سرمایه‌گذاری تأثیر نگذاشته است. یکی از عواملی که می‌تواند در تفکیک توصیه‌های سرمایه‌گذاری به مشتریان کمک کند، خوشه‌بندی آن‌ها به طبقات شخصیتی مختلف و ارائه توصیه متناسب با هر طبقه به هر یک از افراد است.

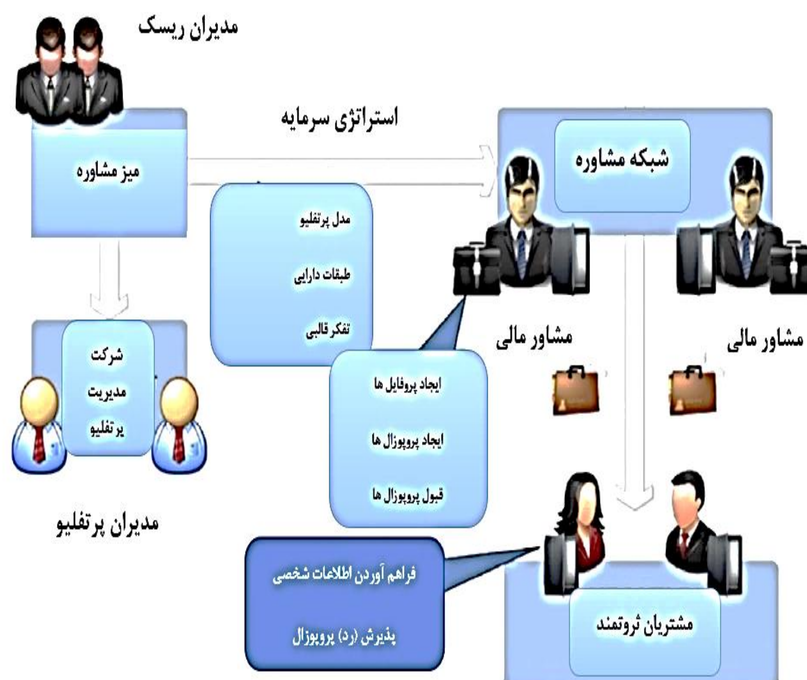
خوشه‌بندی یکی از روش‌های معمول تجزیه‌وتحلیل داده‌های آماری و همچنین یکی از بهترین رویکردهای داده‌کاوی است. مسئله خوشه‌بندی شامل مجموعه‌ای از اشیاء و مجموعه‌ای از ویژگی‌های مرتبط با آن اشیاء است. هدف این است که با استفاده از ویژگی‌هایی

[۱] Robo-advisors
[۲] Statman

که شباهت‌های درون‌گروهی را حداکثر و بین گروهی را کاهش می‌نماید اشیاء را به گروه‌هایی (که خوشه نامیده می‌شود) جدا نماییم. این روش یکی از روش‌های یادگیری بدون نظارت است، زیرا نمونه‌ای برای کمک به حل این الگوریتم در حل این کار ارائه نمی‌دهیم. از الگوریتم‌های خوشه‌بندی در مسائل مختلفی می‌توان استفاده نمود که یکی از کاربردهای آن می‌تواند برای خوشه‌بندی سرمایه‌گذاران با توجه به سطح ریسک آن‌ها باشد.



در **شکل ۱** فرایند یک سیستم مشاوره توصیه‌کننده ترسیم‌شده است. در این فرایند مشتریان بر اساس برخی ویژگی‌ها (مانند نگرش نسبت به ریسک، تخصص سرمایه‌گذاری، افق سرمایه‌گذاری) به گروه‌هایی طبقه‌بندی می‌شوند. درنهایت بر اساس پروفایل شخصیتی هر فرد، یک سبد سرمایه‌گذاری به اون پیشنهاد می‌شود.



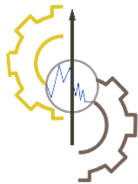
شکل ۱- فرایند کلی توصیه سرمایه‌گذاری (موستو و همکاران[1]، ۲۰۱۵).
Figure 1- The overall investment recommendation process (Musto et al., 2015).

یکی از الزاماتی که برای این فرایند توصیه شخصی نیاز است، این است که ما بتوانیم مشتریان را بر اساس تیپ‌های شخصیتی مختلف به گروه‌هایی دسته‌بندی نماییم. یکی از ابزارهایی که از آن می‌توان برای طبقه‌بندی مشتریان استفاده نمود، خوشه‌بندی است.

خوشه‌بندی سرمایه‌گذاران بر اساس ویژگی‌های مشابه می‌تواند منجر به‌سادگی و استانداردسازی درروند مدیریت سرمایه‌گذاری شود. این الگوریتم‌ها می‌توانند سرمایه‌گذاران را بر اساس عوامل مختلفی مانند سن، درآمد و تحمل ریسک گروه‌بندی کنند. همچنین می‌تواند به مدیران سرمایه‌گذاری کمک کند که گروه‌های متمایزی را در پایگاه سرمایه‌گذاران خود شناسایی کند. بعلاوه، با استفاده از این تکنیک‌ها، مدیران می‌توانند از هرگونه تورشی که منجر به انتخاب نامساعد در تصمیم‌گیری شود جلوگیری نماید. عواملی که از طریق خوشه‌بندی تجزیه‌وتحلیل می‌شوند، می‌توانند تأثیر زیادی در تخصیص و تعادل مجدد دارایی داشته باشند و آن را به ابزاری ارزشمند برای مدیریت سریع‌تر و مؤثر سرمایه‌گذاری تبدیل کنند.

[1] Musto et al.

در این پژوهش ما از روش خوشه‌بندی برای شناسایی انواع مختلف سرمایه‌گذاران استفاده خواهیم کرد. هدف از این پژوهش ایجاد خوشه‌بندی برای گروه‌بندی افراد یا سرمایه‌گذاران بر اساس پارامترهای مربوط به توانایی و تمایل به ریسک است. ما برای دستیابی به این هدف به استفاده از ویژگی‌های مشترک جمعیتی و مالی تمرکز می‌نماییم.



ازآنجایی‌که اطلاعات سرمایه‌گذاران محرمانه است، امکان استفاده از داده‌های ایرانی در خصوص میزان درآمد، ثروت و سایر موارد موردنیاز برای خوشه‌بندی سرمایه‌گذاران وجود نداشت. همچنین با توجه به محدودیت‌های قانونی سازمان بورس مبنی بر عدم افشا اطلاعات مشتریان و محرمانه بودن آن‌ها، به‌ناچار در این پژوهش از داده‌های آمریکا استفاده‌شده است؛ بنابراین داده‌های مورداستفاده برای این پژوهش مربوط به پرسشنامه تأمین مالی مصرف‌کننده[1] *(SCF)* است که توسط شوارى فدرال رزو انجام‌شده است[2]. این نظرسنجی شامل پاسخ‌هایی در مورد آمارهای جمعیتی خانوار، خالص دارایی‌های مالی و غیرمالی در سال ۲۰۰۷ (قبل بحران) و ۲۰۰۹ (بعد از بحران) است. در ادامه به بررسی ادبیات نظری و پیشینه پژوهش‌های انجام‌شده در خصوص خوشه‌بندی پرداخته و پس از آن در خصوص روش انجام پژوهش، داده‌های مورداستفاده و روش‌های خوشه‌بندی مورداستفاده می‌پردازیم. درنهایت در بخش آخر به تحلیل نتایج حاصل از خوشه‌بندی پرداخته‌شده است و مدل بومی‌ای برای ایران ارائه می‌شود.

۲- مبانی نظری و پیشینه پژوهش

داده‌کاوی به‌عنوان علم تحلیل داده‌ها و باهدف دست یافتن به بینش و دانشی نسبت به داده‌های موردمطالعه، معرفی می‌شود. پژوهشگران در اکثر حوزه‌های علمی اعم از مدیریت، اقتصاد، علوم پزشکی، مهندسی و زیست‌شناسی با بحران رشد سریع اطلاعات و وجود داده‌هایی با ابعاد بالا مواجه هستند که با بهره‌گیری از این روش، درصدد درک روابط بین پدیده‌های موجود می‌باشند (ویلیامز و همکاران[3]، ۲۰۱۰). خوشه‌بندی به‌عنوان یکی از روش‌های مهم داده‌کاوی به‌منظور استخراج اطلاعات مفید از مجموعه‌های مختلف داده در ابعاد بالا معرفی می‌شود. در سال‌های اخیر مسئله خوشه‌بندی و کشف روابط و روندهای موجود در سری‌های زمانی، موردتوجه بسیاری از محققان در حوزه‌های مختلف علوم ازجمله پژوهش‌های مالی قرارگرفته است. خوشه‌بندی سری‌های زمانی امکان تجزیه‌وتحلیل حجم زیادی از داده‌ها را در مقاطع زمانی مختلف، فراهم می‌کند (چاندرا و همکاران[4]، ۲۰۰۸). به‌طورکلی خوشه‌بندی یکی از روش‌های داده‌کاوی است که در آن، داده‌های مشابه در گروه‌های مرتبط یا همگن قرار می‌گیرند (رای و سینگ[5]، ۲۰۱۰). فرایند خوشه‌بندی شامل بخش‌بندی N داده در k گروه یا خوشه است. اگر مجموعه‌ای از داده‌ها به‌صورت $X=\{x_1,x_2,x_3,\dots,x_N\}$ در نظر گرفته شود، داده‌هایی که دارای بیشترین شباهت به یکدیگر هستند در یک خوشه و داده‌هایی که شباهت کمتری دارند در دیگر خوشه‌ها طبقه‌بندی می‌شوند. بسته به اینکه از کدام روش خوشه‌بندی استفاده می‌شود ترکیب خوشه‌ها می‌تواند متفاوت باشد (ازوگیو و همکاران[6]، ۲۰۲۱). خوشه‌بندی داده‌های سری زمانی معمولاً برای کشف الگوهای موجود در مجموعه سری‌های زمانی مورداستفاده قرار می‌گیرد (وانگ و همکاران[7]، ۲۰۰۲). این وظیفه خود به دو بخش مجزا تقسیم می‌شود، بخش اول شامل یافتن الگوهایی است که به‌صورت مکرر در مجموعه‌های سری زمانی رخ می‌دهد (فیو و همکاران[8]، ۲۰۰۱؛ چیو و همکاران[9]، ۲۰۰۳) و بخش دوم روش‌هایی است که به کشف الگوهایی که به‌ندرت در سری‌های زمانی رخ می‌دهند می‌پردازد و همچنین اتفاقاتی که دارای تأثیرات شگفت‌آوری بر روند سری‌های زمانی هستند را بررسی می‌کند (کوف و همکاران[10]، ۲۰۰۲؛ لنگ و همکاران[11]، ۲۰۰۹). به‌طورکلی تجزیه‌وتحلیل سری‌های زمانی به سه هدف انجام می‌شود (پالز و همکاران[12]، ۲۰۰۳):

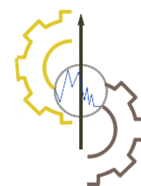
---

[1] Survey of Consumer Finances
[2] https://www.federalreserve.gov/econres/scf_2009p.htm
[3] Williams et al.
[4] Chandra et al.
[5] Rai and Singh
[6] Ezugwu et al.
[7] Wang et al.
[8] Fu et al.
[9] Chiu et al.
[10] Keogh et al.
[11] Leng et al.
[12] Polz et al.

– شناخت تغییرات پویا در سری‌های زمانی: تشخیص همبستگی بین سری‌های زمانی، به‌عنوان‌مثال: در پایگاه داده‌های مالی می‌توان از آن برای یافتن شرکت‌هایی با حرکت مشابه سهام استفاده کرد (هی و همکاران[1]، ۲۰۱۲).

– پیش‌بینی و توصیه: یک روش ترکیبی از ادغام خوشه‌بندی و تقریب عملکرد در هر خوشه می‌تواند به تحلیلگران در پیش‌بینی و توصیه کمک کند به‌عنوان‌مثال، در پایگاه‌های اطلاعاتی علمی، می‌تواند در حل مشکلاتی مانند یافتن الگوهای باد مغناطیسی خورشیدی برای پیش‌بینی الگوی امروز کمک کند (اسفتسوس و سیریوپولوس[2]، ۲۰۰۴؛ گریو و پیریسز[3]، ۲۰۱۰).

– کشف الگو: باهدف کشف الگوهای جالب در پایگاه داده. به‌عنوان‌مثال، در پایگاه داده بازاریابی، می‌توان الگوهای مختلف فروش روزانه یک محصول خاص را در یک فروشگاه کشف کرد (آقابزرگی و همکاران[4]، ۲۰۱۵).



خوشه‌بندی به دودسته بخشی (جزء بندی) و سلسله مراتبی تقسیم‌بندی می‌شود که در ادامه به تعریف و دسته‌بندی هرکدام می‌پردازیم.

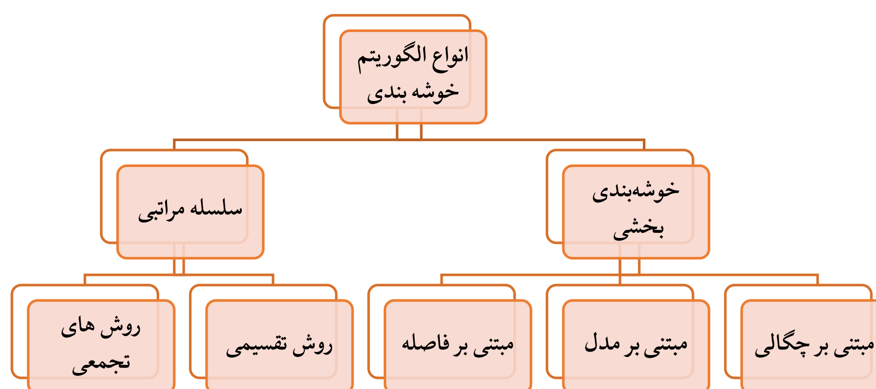
۲-۱- خوشه‌بندی بخشی (جزء بندی)

تقسیم مجموعه داده‌ها به زیرمجموعه‌های بدون همپوشانی به‌گونه‌ای که هر داده دقیقاً در یک زیرمجموعه قرار داشته باشد.خوشه‌بندی سلسله مراتبی به دودسته تقسیم می‌شود:

روش‌های تجمعی (Agglomerative) (روش پایین به بالا). با هر داده در یک خوشه شروع می‌کند. به‌طوری تکراری، در هر مرحله خوشه‌های نزدیک به هم را ترکیب می‌کند تا درنهایت یک خوشه باقی بماند.

روش‌های تقسیمی (Divisible) (روش بالا به پایین). با کل داده‌ها به‌عنوان یک خوشه شروع می‌کند. به‌طور تکراری، داده‌ها را دریکی از خوشه‌ها تقسیم می‌کند تا هنگامی‌که فقط یک داده در هر خوشه باشد.

همچنین می‌توان رویکردهای کلی الگوریتم‌های خوشه‌بندی را به‌صورت **شکل ۲** ارائه نمود:



شکل ۲- انواع الگوریتم‌های خوشه‌بندی (ساکسنا و همکاران[5]، ۲۰۱۷).

Figure 2- Types of clustering algorithms (Saxena et al., 2017).

انواع مختلفی از تکنیک‌های خوشه‌بندی وجود دارد که با توجه به استراتژی، شناسایی گروه‌بندی‌ها متفاوت هستند. انتخاب اینکه کدام تکنیک باید اعمال شود به ماهیت و ساختار داده بستگی دارد. در این بخش، به دو روش خوشه‌بندی پرداخته خواهد شد.

---

[1] He et al.
[2] Sfetsos and Siriopoulos
[3] Graves and Pedrycz
[4] Aghabozorgi et al.
[5] Saxena et al.

نوراحمدی و صادقی / مدیریت نوآوری و راهبردهای عملیاتی، دوره ۲، شماره ۲، تابستان ۱۴۰۰، صفحه: ۱۹۴-۱۸۰

۲-۲- خوشه‌بندی k-means

k-means معروف‌ترین روش خوشه‌بندی توسط مک کوئین[1] (۱۹۶۷) پیشنهاد شد؛ این روش به‌عنوان یک الگوریتم خوشه‌بندی کلاسیک در پژوهش‌های علمی و کاربردهای صنعتی مورداستفاده قرار می‌گیرد. هدف الگوریتم *k-means* یافتن و گروه‌بندی نقاط داده در کلاس‌هایی که شباهت زیادی بین آن‌ها وجود دارد این شباهت به‌عنوان نقطه مقابل فاصله بین داده‌ها درک می‌شود درواقع هرچه نقاط داده نزدیک‌تر باشند، احتمال تعلق آن‌ها به یک خوشه بیشتر است. ایده اصلی این الگوریتم این است که n شیء داده را به n خوشه تقسیم کند به‌نحوی‌که مجموع مربعات نقاط داده در هر خوشه به مرکز خوشه کوچک‌ترین باشد (تدوریدیس و کرتی[2]، ۲۰۰۰). این الگوریتم مرکز k را پیدا می‌کند و هر نقطه داده را دقیقاً به یک خوشه و باهدف به حداقل رساندن واریانس درون خوشه (که اینرسی نامیده می‌شود) اختصاص می‌دهد. این روش معمولاً از فاصله اقلیدسی (فاصله معمولی بین دونقطه) استفاده می‌کند، اما سایر معیارهای فاصله می‌توانند مورداستفاده قرار گیرند. الگوریتم *k-means* یک بهینه محلی برای یک k داده‌شده ارائه می‌دهد و به‌صورت زیر عمل می‌کند:

- این الگوریتم تعداد خوشه‌ها را مشخص می‌کند.
- نقاط داده به‌طور تصادفی به‌عنوان مراکز خوشه انتخاب می‌شوند.
- هر نقطه داده به مرکز خوشه که نزدیک‌ترین نقطه به آن است اختصاص داده می‌شود.
- مراکز خوشه به میانگین تخصیص به‌روزرسانی می‌شوند.
- گام ۳ تا ۴ تکرار می‌شوند تا زمانی که تمام مراکز خوشه بدون تغییر بمانند.

به عبارت ساده‌تر، به‌طور تصادفی در اطراف تعداد مشخصی از مراکز جرم در هر تکرار حرکت کرده و هر نقطه داده به نزدیک‌ترین مرکز جرم اختصاص داده می‌شود (تاستات و همکاران[3]، ۲۰۲۰).

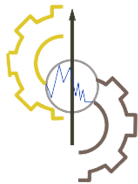
۲-۳- روش خوشه‌بندی انتشار وابستگی

انتشار وابستگی[4] با ارسال پیام بین نقاط داده تا همگرایی، خوشه‌ها را ایجاد می‌کند. برخلاف الگوریتم‌های خوشه‌بندی مانند *k-means*، انتشار وابستگی به تعداد خوشه‌هایی که باید قبل از اجرای الگوریتم تعیین یا تخمین زده شوند نیاز ندارد. دو پارامتر مهم در انتشار وابستگی که برای تعیین تعداد خوشه‌ها از استفاده می‌شوند شامل: اولویت که تعداد نمونه‌ها (نمونه‌های اولیه) را کنترل می‌کند استفاده می‌شود؛ و عامل میرایی که مسئولیت و در دسترس بودن پیام‌ها را برای جلوگیری از نوسانات عددی در هنگام به‌روزرسانی این پیام‌ها کاهش می‌دهد. یک مجموعه داده با استفاده از تعداد کمی نمونه توصیف می‌شود. این‌ها اعضای مجموعه ورودی هستند که نمایانگر خوشه‌ها هستند. الگوریتم انتشار وابستگی، مجموعه‌ای از شباهت‌های دوبه‌دو بین نقاط داده را در برمی‌گیرد و با به حداکثر رساندن شباهت کل بین نقاط داده و نمونه‌های آن‌ها خوشه‌ها را پیدا می‌کند. پیام‌های ارسال‌شده بین جفت‌ها، نشان‌دهنده‌ی مناسب بودن یک نمونه به‌عنوان نمونه‌ای از دیگری است که در پاسخ به مقادیر سایر جفت‌ها به‌روزرسانی می‌شود. این به‌روزرسانی به‌صورت تکراری رخ می‌دهد تا همگرایی حاصل شود که در آن نقطه نمونه‌های نهایی انتخاب می‌شوند و خوشه‌بندی نهایی به دست می‌آید (تاستات و همکاران، ۲۰۲۰). روش خوشه‌بندی به‌طور گسترده‌ای در علوم مختلف و برای مقاصد گوناگونی به کار گرفته می‌شود که در **جدول ۱** به بررسی پیشینه پژوهش‌های انجام‌شده به‌طور خاص در خصوص کاربرد روش‌های خوشه‌بندی در پرتفوی و مدیریت ریسک پرداخته‌شده است.

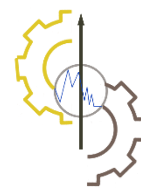
۳- روش پژوهش

همان‌گونه که در مقدمه مطرح شد، ازآنجایی‌که اطلاعات سرمایه‌گذاران محرمانه است، برای انجام این پژوهش امکان استفاده از داده‌های ایرانی در خصوص میزان درآمد، ثروت و سایر موارد موردنیاز برای خوشه‌بندی سرمایه‌گذاران وجود نداشت. همچنین با توجه به

[1] MacQueen
[2] Thuraisingham and Ceruti
[3] Tatsat et al.
[4] Affinity Propagation Clustering



محدودیت‌های قانونی سازمان بورس مبنی بر عدم افشا اطلاعات مشتریان و محرمانه بودن آن‌ها، به‌ناچار در این پژوهش از داده‌های مربوط به پرسشنامه تأمین مالی مصرف‌کننده *(SCF)* است که توسط شوارى فدرال رزو انجام‌شده است[۱]، استفاده‌شده است. مجموعه داده‌ها شامل ۳۸۶۶ مشاهده با ۱۲ ویژگی است. کلیه محاسبات انجام‌شده در این پژوهش توسط پایتون ۳/۸ انجام‌شده است.



مراحل آماده‌سازی متغیرهای پیش‌بینی‌شده به شرح زیر است:

الف) دارایی‌های ریسکی و بدون ریسک را برای همه افراد در داده‌های نظرسنجی محاسبه می‌نماییم. دارایی‌های ریسکی و بدون ریسک به شرح زیر تعریف می‌شوند.

- دارایی‌های ریسکی: شامل سرمایه‌گذاری در صندوق‌های سرمایه‌گذاری مشترک سهام، اوراق قرضه، کالاها و املاک و مستغلات و سرمایه انسانی است.
- دارایی‌های بدون ریسک: مانده‌حساب تراز، پس‌انداز، گواهی سپرده و سایر مانده‌های نقدی و معادل آن.

ب) نسبت دارایی‌های ریسکی به‌کل دارایی‌های سرمایه‌گذار را محاسبه نموده و به‌عنوان معیار تحمل ریسک در نظر می‌گیریم.

در ادبیات مالی، سرمایه‌گذار باهوش کسی است که در هنگام تغییر در بازار میزان تحمل ریسک خود را تغییر ندهد؛ بنابراین سرمایه‌گذارانی که در طول دوره زمانی کمتر از ۱۰% تحمل ریسک خود را تغییر می‌دهند، به‌عنوان سرمایه‌گذاران باهوش در نظر گرفته می‌شود. البته این‌یک معیار کیفی است و قابل‌تغییر است (تاستات و همکاران، ۲۰۲۰).

جدول ۱- کاربردهای خوشه‌بندی در بهینه‌سازی پرتفوی و مدیریت ریسک.
Table 1- Applications of clustering in portfolio optimization and risk management.

عنوان مقاله	نویسنده (گان)	نتایج
برابری ریسک سلسله مراتبی: حسابداری وابستگی‌های ناشی از تخصیص چندعاملی چند دارایی.	(لوهر و همکاران[۲]، ۲۰۲۰)	بررسی استراتژی‌های متنوع سازی بر اساس خوشه‌بندی سلسله مراتبی
آیا پرتفوی مبتنی بر روش‌های یادگیری ماشینی می‌تواند از پرتفوی مبتنی بر ریسک سنتی بهتر عمل کنند؟	(جین و جین[۳]، ۲۰۱۹)	در این مقاله تأثیر انتخاب غلط ماتریس کوواریانس را بر عملکرد روش‌های مختلف تخصیص بررسی نموده و سپس بررسی می‌کند که آیا عملکرد HRP مبتنی بر روش‌های یادگیری ماشین بهتر از پرتفوی مبتنی بر روش‌های تعدیل ریسک سنتی است یا نه؟
ریسک برابر سلسله مراتبی برای ساخت پرتفوی	(رافینوت[۴]، ۲۰۱۸)	نتایج نشان می‌دهد که پرتفوی HERC[۵] مبتنی بر معیارهای ریسک نزولی، ازنظر آماری عملکرد بهتری نسبت به معیارهای CDaR برای تعدیل ریسک به دست می‌آورند.
استراتژی پرتفوی پویا با استفاده از رویکرد خوشه‌بندی	(رن و همکاران[۶]، ۲۰۱۷)	استراتژی پرتفوی پویا مبتنی بر ساختارهای متغیر شبکه‌های MST در بازار سهام چنین ارائه می‌شود.
الگوریتم‌های خوشه‌بندی برای ساخت پرتفوی ریسک تعدیل‌شده	(لئون و همکاران[۷]، ۲۰۱۷)	در این مقاله عملکرد هفت پرتفوی ایجادشده با استفاده از تکنیک‌های تجزیه‌وتحلیل خوشه‌بندی برای طبقه‌بندی دارایی‌ها به دسته‌ها و سپس استفاده از بهینه‌سازی کلاسیک در داخل هر خوشه برای انتخاب بهترین دارایی‌ها در داخل هر طبقه را بررسی می‌نماید.

[۱] https://www.federalreserve.gov/econres/scf_2009p.htm
[۲] Lohre et al.
[۳] Jain and Jain
[۴] Raffinot
[۵] Hierarchical Equal Risk Contribution (HERC)
[۶] Ren et al.
[۷] León et al.

نوراحمدی و صادقی / مدیریت نوآوری و راهبردهای عملیاتی، دوره ۲، شماره ۲، تابستان ۱۴۰۰، صفحه: ۱۹۴-۱۸۰

جدول ۱- ادامه.
Table 1- Continued.

عنوان مقاله	نویسنده (گان)	نتایج
انتخاب دارایی مبتنی بر خوشه‌بندی سلسله مراتبی	(رافینوت، ۲۰۱۷)	این مقاله یک روش تخصیص دارایی مبتنی بر خوشه‌بندی سلسله مراتبی پیشنهاد می‌نماید که از تئوری شبکه و تکنیک‌های یادگیری ماشین استفاده می‌نماید. نتایج تجربی آن‌ها نشان می‌دهد که پرتفوی مبتنی بر خوشه‌بندی سلسله مراتبی پایدار و واقعاً متنوع هستند و نسبت به تکنیک‌های معمول بهینه‌سازی ازنظر آماری عملکرد بهتری را برای تعدیل ریسک به دست می‌آورند.
خوشه‌بندی سری‌های زمانی مالی در شرایط ریسکی	(دیورانت و همکاران[۱]، ۲۰۱۴)	در این پژوهش روشی برای خوشه‌بندی سری‌های زمانی مالی با توجه به پیوستگی توزیع آن‌ها ارائه‌شده است این روش بر اساس محاسبه ضرایب همبستگی اسپیرمن استخراج‌شده است. نتایج این تحقیق می‌تواند برای ایجاد پرتفوی مالی با حداقل ریسک استفاده شود
تحلیل خوشه در بهینه‌سازی پرتفوی	(تولا و همکاران[۲]، ۲۰۰۸)	در این پژوهش مسئله‌ی عدم اطمینان آماری ماتریس همبستگی را در بهینه‌سازی پرتفوی در نظر می‌گیرد. این پژوهش نشان می‌دهد که با فرض شرایط ایده آل با استفاده از الگوریتم‌های خوشه‌بندی می‌توان قابلیت اطمینان پرتفوی‌ها را ازنظر نسبت ریسک پیش‌بینی‌شده و ریسک تحقق‌یافته بهبود بخشید.
خوشه‌بندی سری‌های زمانی مالی با کاربرد آن در افزایش ردیابی شاخص پرتفوی	(داز و کینکاتی[۳]، ۲۰۰۵)	نتایج مبتنی بر خوشه‌بندی با نتایج تکنیک‌های تصادفی مقایسه می‌شود و اهمیت خوشه‌بندی را در کاهش نویز و پایداری روش‌های پیش‌بینی به‌ویژه برای ردیابی شاخص پیشرفته نشان می‌دهد.

ابتدا میزان تحمل ریسک را برای هر یک از سال‌های ۲۰۰۷ و ۲۰۰۹ محاسبه می‌نماییم و سپس میانگین آن را به‌عنوان معیار تحمل ریسک در نظر می‌گیریم.

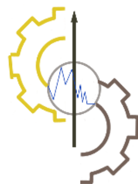
در مرحله بعدی ویژگی‌ها را کاهش می‌دهیم. در **جدول ۲** ویژگی‌های موردنیاز برای خوشه‌بندی ارائه‌شده است.

جدول ۲- ویژگی‌های موردنیاز برای خوشه‌بندی (تاستات و همکاران، ۲۰۲۰).
Table 2- Properties required for clustering (Tastat et al.,2020).

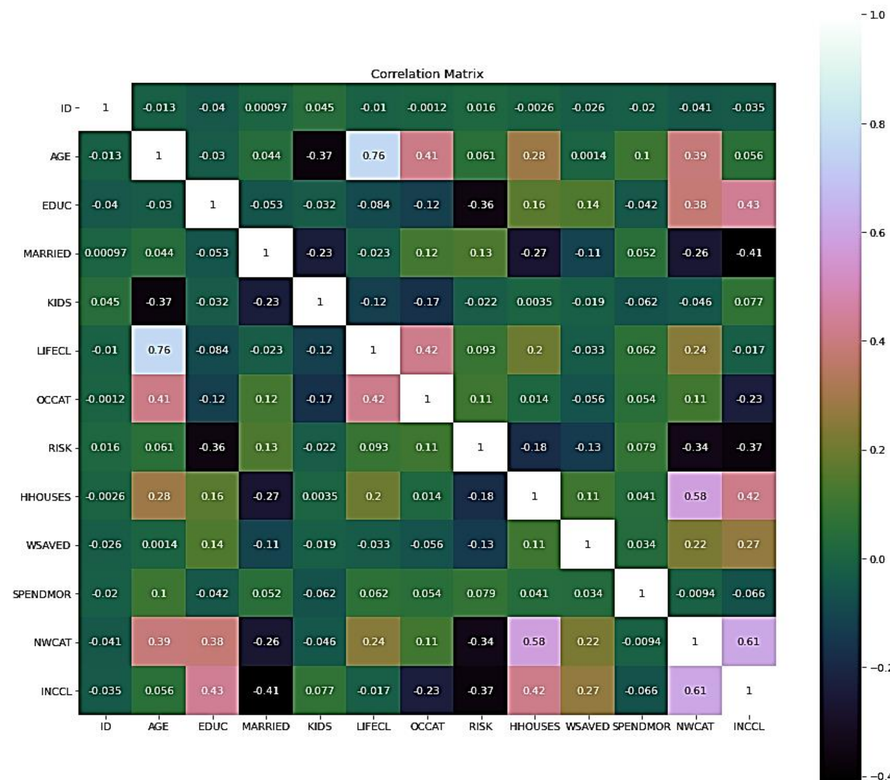
خصوصیات جمعیت شناختی	خصوصیات مالی	خصوصیات رفتاری
سن AGE	مالک خانه HHOUSES	تمایل به پذیرش ریسک RISK
تحصیلات EDUC	هزینه ها در مقابل درآمد SPENDMOR	تمایل به هزینه کردن SPENDMOR
تاهل MARRIED	ثروت خالص NWCAT	
تعداد فرزندان KIDS	درآمد INCCL	
مرحله زندگی LIFECYCL		
اشتغال OCCU		

یکی از روش‌های متداول برای مصورسازی داده‌ها استفاده از طریق نقشه‌های حرارتی[۴] است. نمودار نقشه‌های حرارتی از یک طیف مقیاس رنگی کدگذاری می‌شود که در آن همبستگی‌های مثبت قوی (همبستگی = یک) سبز تیره، دارایی‌های غیر همبسته (همبستگی = صفر) و دارایی‌های دارای همبستگی منفی (همبستگی= منفی یک) را به رنگ آبی نشان می‌دهد. نقشه‌های حرارتی مفید هستند اما بااین‌حال، آن‌ها فقط می‌توانند یک بعد از اطلاعات (میزان همبستگی بین دو ویژگی) را انتقال بدهند.

[۱] Durante et al.
[۲] Tola et al.
[۳] Dose and Cincotti
[۴] Heatmap



در **شکل ۳** میزان همبستگی بین این ویژگی‌ها نشان داده می‌شود:



شکل ۳- نقشه حرارتی همبستگی بین ویژگی‌ها.

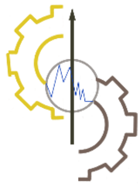
Figure 3- Thermal map of correlation between properties.

تکنیک‌های مختلفی برای خوشه‌بندی وجود دارد و آن‌ها ازنظر استراتژی شناسایی گروه‌بندی متفاوت هستند. انتخاب تکنیک مورداستفاده به ماهیت و ساختار داده بستگی دارد. در این پژوهش ما از روش خوشه‌بندی *k-means* و انتشار وابستگی استفاده نموده‌ایم.

همان‌گونه که در ادبیات نظری توضیح دادیم، *k-means* یکی از معروف‌ترین تکنیک‌های خوشه‌بندی است. الگوریتم *k-means* برای پیدا کردن و گروه‌بندی نقاط داده در کلاس‌هایی است که شباهت زیادی بین آن‌ها وجود دارد. این شباهت برخلاف فاصله بین نقاط بین داده است. هرچه نقاط داده نزدیک‌تر باشند، احتمال تعلق آن‌ها به خوشه یکسان بیشتر است.

در الگوریتم خوشه‌بندی *k-Means* تعداد خوشه‌ها پیش از خوشه‌بندی تعیین‌شده و با پارامتر *k* به‌عنوان ورودی به الگوریتم وارد می‌شود. توضیح این الگوریتم به این صورت است که در ابتدا از بین داده‌های موردبررسی (سری زمانی) *k* سری زمانی را به‌صورت کاملاً تصادفی انتخاب و هرکدام را در یک خوشه‌ی مجزا قرار داده و آن‌ها را به‌عنوان مراکز ثقل خوشه‌ها در نظر می‌گیریم؛ که در این پژوهش $k=7$ (تعداد خوشه بهینه) به‌وسیله روش *Sum of Squared Errors (SSE)* و *Silhouette score* تعیین می‌شود (نتایج در **شکل ۲** نمایش داده‌شده است). در این مرحله به انتصاب *n-k* داده‌ی سری زمانی باقی‌مانده به خوشه‌های تشکیل‌شده در مرحله‌ی قبل می‌پردازیم. این انتصاب با معیار کمترین فاصله‌ی هر سری زمانی با مرکز ثقل خوشه‌ها صورت می‌گیرد؛ بدین‌صورت که با بهره‌گیری از یک معیار اندازه‌گیری فاصله، فاصله‌ی هر سری زمانی از مراکز ثقل خوشه‌های مختلف را با یکدیگر مقایسه کرده و با معیار کمترین فاصله، سری‌های زمانی را به خوشه‌ها اختصاص می‌دهیم. در این مرحله لازم است با توجه به اعضای هر خوشه مرکز ثقل جدیدی برای خوشه‌ها تعیین شود. در این روش میانگین داده‌های موجود در هر خوشه به‌عنوان مرکز ثقل جدید آن خوشه در نظر گرفته می‌شود. درنتیجه به‌منظور تعیین مرکز ثقل خوشه‌ها به میانگین‌گیری از داده‌های سری زمانی موجود در هر خوشه می‌پردازیم پس از تعیین مرکز ثقل جدید برای هر خوشه لازم است مجدداً دو مرحله قبل را تکرار کنیم این مراحل تا زمانی که دیگر مراکز ثقل به‌دست‌آمده در مرحله‌ی آخر تغییر نکند، ادامه می‌یابد.

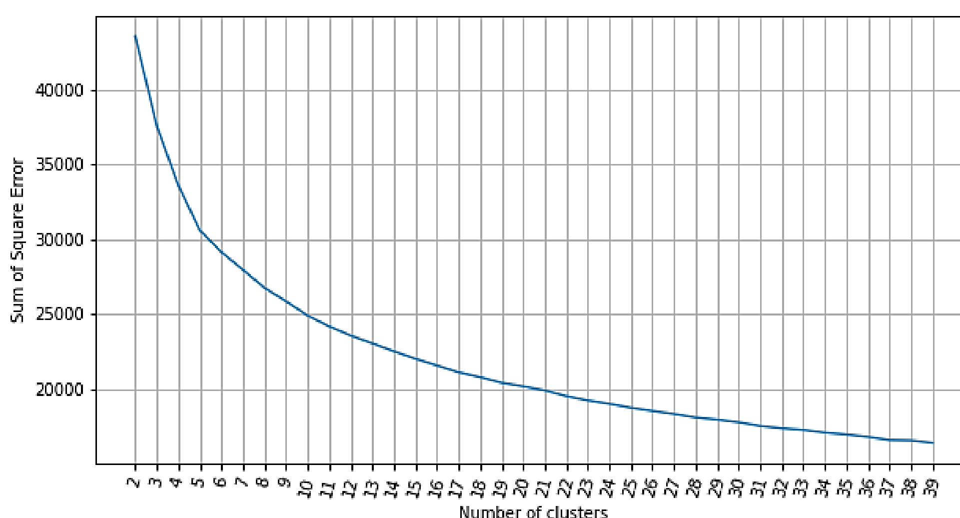
روش دیگری که برای خوشه‌بندی استفاده کرده‌ایم، روش انتشار وابستگی[1] است. انتشار وابستگی با ارسال پیام بین نقاط داده تا همگرایی، خوشه‌ها را ایجاد می‌کند. برخلاف الگوریتم‌های خوشه‌بندی مانند *k-means*، انتشار وابستگی به تعداد خوشه‌هایی که باید قبل از اجرای الگوریتم تعیین یا تخمین زده شوند نیاز ندارد (تاستات و همکاران، ۲۰۲۰).



تجزیه‌وتحلیل داده‌ها به‌منظور پاسخ‌گویی به سؤالات پژوهشی اهمیت فراوانی داشته، ازاین‌رو در تمام پژوهش‌ها از اصلی‌ترین و مهم‌ترین بخش‌های تحقیق به شمار می‌رود. لذا می‌توان از تجزیه‌وتحلیل اطلاعات به‌عنوان جزء لاینفک پژوهش‌های علمی، به‌منظور پاسخ‌گویی به مسائل تدوین‌شده و همچنین تائید یا رد فرضیه‌های پژوهش نام برد. در ابتدا به این مسئله می‌پردازیم که از بین روش‌های مختلف خوشه‌بندی کدام‌یک بهترین روش است. ضریب *silhouette* در *sklearn.metrics.sil houette_score* را می‌توان به‌منظور ارزیابی نتایج خوشه‌بندی استفاده نمود. ضریب *silhouette* بالاتر حاکی از مدلی با خوشه‌های بهتر است. بدین منظور از کتابخانه *sklearn* و با استفاده از معیار *metrics* در پایتون به ارزیابی دو روش خوشه‌بندی ذکرشده پرداخته می‌شود. نتایج ارزیابی به شرح زیر است:

ضریب *k-means* برابر با ۰/۱۷ و انتشار وابستگی برابر با ۰/۰۹۷ شد؛ بنابراین ما روش *k-means* را به‌عنوان روش خوشه‌بندی مطلوب انتخاب می‌نماییم.

به منظور ارزیابی مدل‌ها، ابتدا با استفاده از روش مجموع مربعات خطا[2] و معیار نیم‌رخ[3] تعداد خوشه بهینه را پیدا می‌کنیم. معیار نیم‌رخ، هم به پیوستگی *(Cohesion)* درون خوشه‌ها و هم به میزان تفکیک‌پذیری آن‌ها بستگی دارد. مقدار نیم‌رخ برای هر نقطه، میزان تعلق آن را به خوشه‌اش در مقایسه با خوشه مجاور اندازه می‌گیرد.



شکل ۴- محاسبه تعداد خوشه بهینه با استفاده از مجموع مربعات خطا.

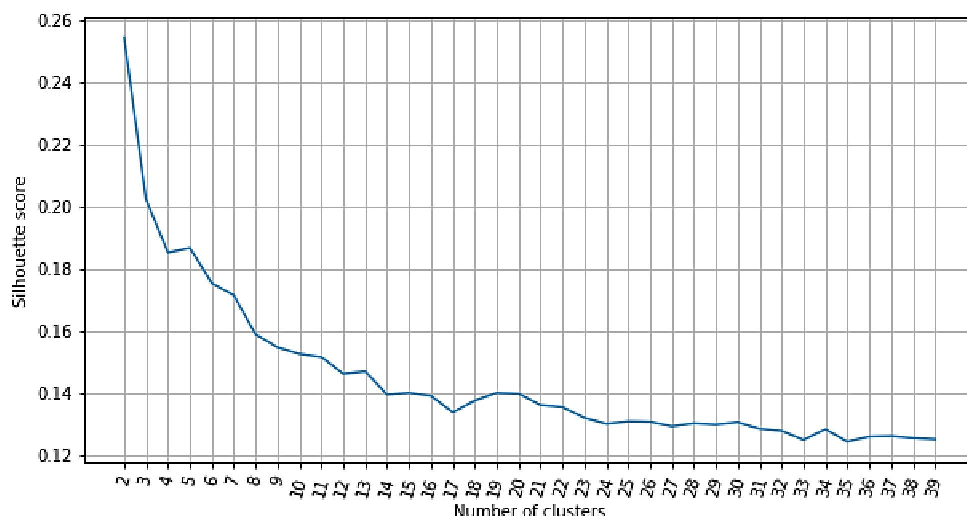
Figure 4- Calculate the number of optimal clusters using the sum of the error squares.

با نگاهی به هر دو **شکل ٤** و **٥** به نظر می‌رسد تعداد بهینه خوشه‌ها حدود ۷ باشد. همان‌گونه که ملاحظه می‌شود، با افزایش تعداد خوشه‌ها از ٦، *SSE* درون خوشه‌ها شروع به کم شدن می‌کند.

۴- تحلیل داده‌ها

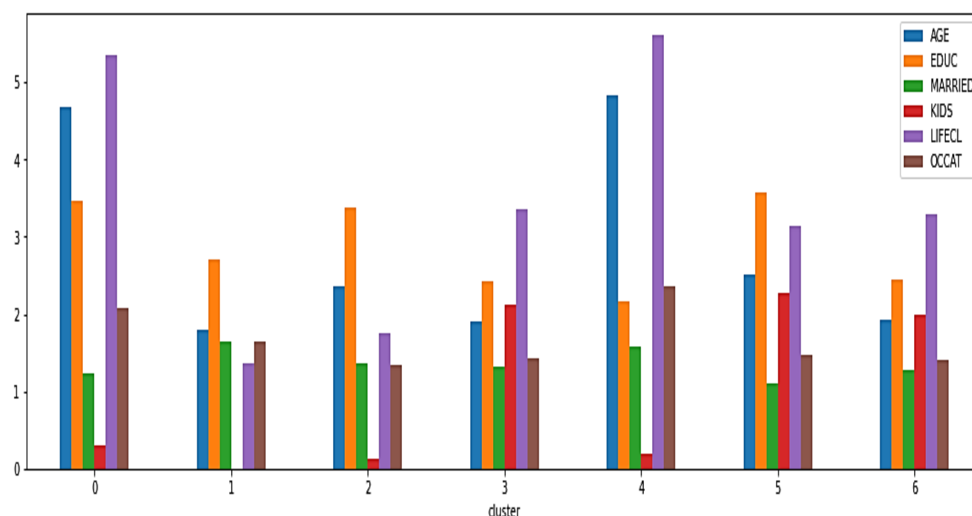
با توجه به اینکه روش *k-means* معیار بهتری نسبت به روش انتشار وابستگی بود، در ادامه به خوشه‌بندی با استفاده از روش *k-means* برای ویژگی‌های جمعیت شناختی و ویژگی‌های رفتاری و مالی می‌پردازیم.

[1] Affinity Propagation
[2] Sum of Squared Errors (SSE)
[3] Silhouette Score



شکل ۵- محاسبه تعداد خوشه بهینه با استفاده از معیار نیم‌رخ.

Figure 5- Calculate the number of optimal clusters using the profile criterion.



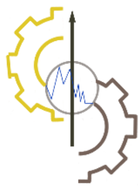
شکل ۶- خوشه‌بندی بر اساس ویژگی‌های جمعیت شناختی.

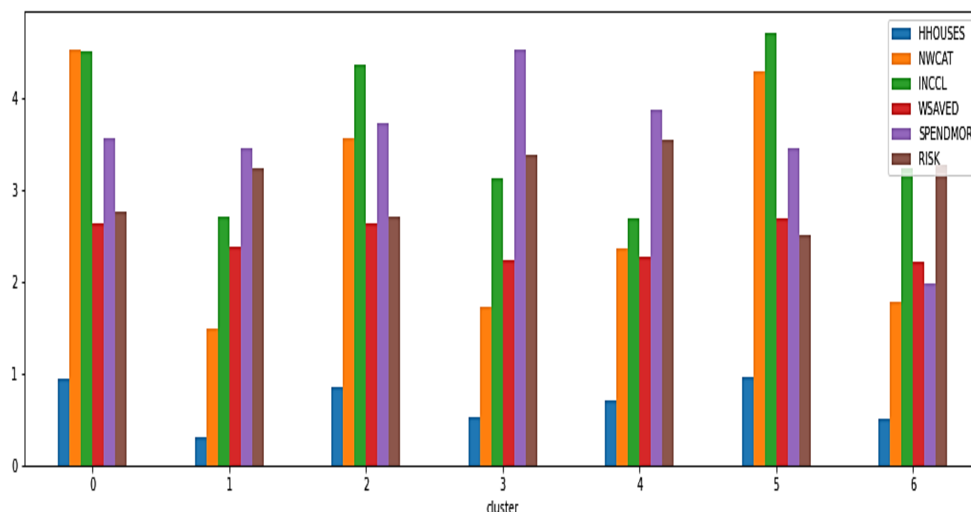
Figure 6- Clustering based on demographic characteristics.

شکل ٦ میانگین مقدار صفات را برای هر یک از خوشه‌ها نشان می‌دهد. به‌عنوان‌مثال، با مقایسه خوشه صفر و یک، خوشه صفر دارای میانگین سنی کمتر و درعین‌حال میانگین تحصیلات بالاتر است. ازنظر ازدواج و تعداد فرزندان، این دو خوشه شبیه هم هستند؛ بنابراین افراد در خوشه صفر بر اساس ویژگی‌های جمعیت شناختی در مقایسه با افراد خوشه یک، تحمل ریسک بیشتری دارند.

در **شکل ۷** میانگین صفات را برای هریک از خوشه‌ها بر روی ویژگی‌های مالی و رفتاری نشان می‌دهد. به‌عنوان‌مثال، در مقایسه خوشه صفر و یک، خوشه صفر دارای متوسط مالکیت خانه بالاتر، متوسط خالص دارایی و درآمد بالاتر و تمایل به خطرپذیری کمتر است. ازنظر پس‌انداز در مقایسه با درآمد و تمایل به پس‌انداز، این دو خوشه قابل‌مقایسه هستند؛ بنابراین می‌توانیم تصور کنیم که افراد موجود در خوشه صفر، در مقایسه با خوشه یک به‌طور متوسط توانایی بالاتری و درعین‌حال تمایل کمتری برای ریسک‌پذیری دارند که در **شکل ۸** نشان داده شده است.

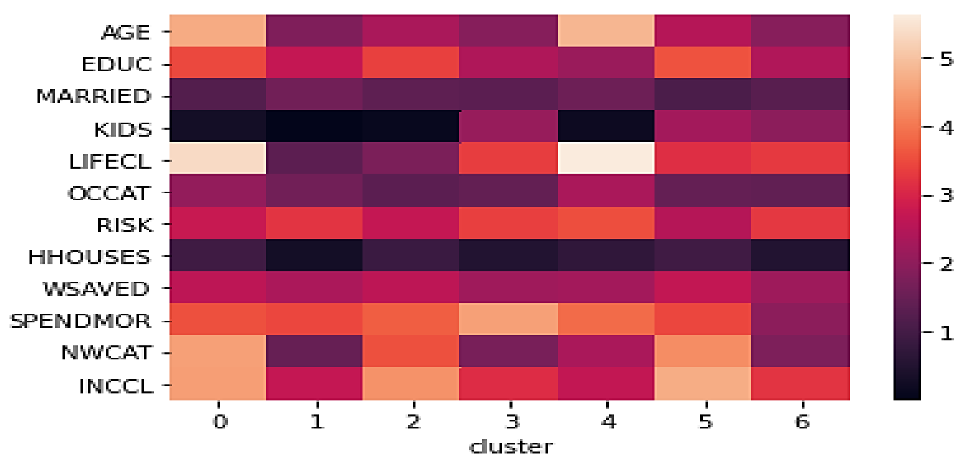
نوراحمدی و صادقی / مدیریت نوآوری و راهبردهای عملیاتی، دوره ۲، شماره ۲، تابستان ۱۴۰۰، صفحه: ۱۹۴-۱۸۰





شکل ۷- خوشه‌بندی بر اساس ویژگی‌های مالی و ویژگی‌های مربوط به تمایل به ریسک.

Figure 7- Clustering based on financial characteristics and risk-related characteristics.



شکل ۸- ترکیب نتایج خوشه‌بندی بر اساس ویژگی‌های جمعیتی، مالی و رفتاری.

Figure 8- Combining clustering results based on demographic, financial and behavioral characteristics.

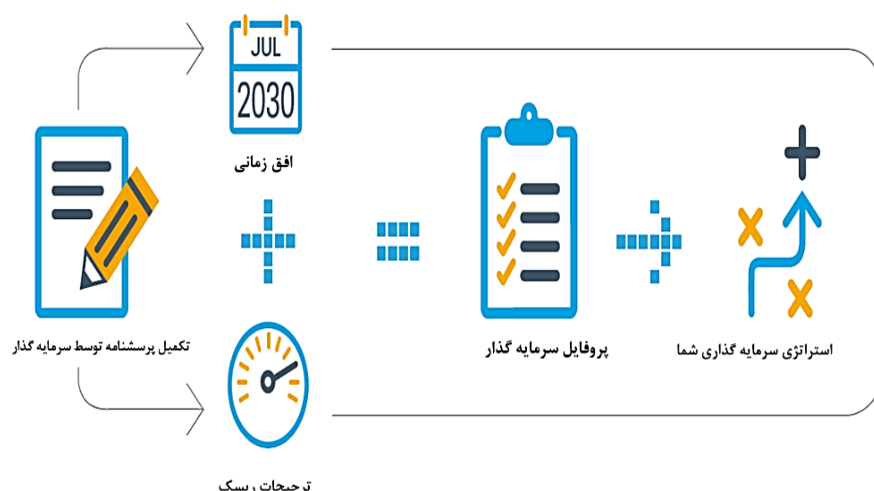
با ترکیب اطلاعات مربوط به ویژگی های جمعیتی، مالی و رفتاری برای خوشه صفر و یک، توانایی کلی برای پذیرش ریسک برای خوشه‌های صفر در مقایسه با خوشه یک بیشتر است. تجزیه‌وتحلیل مشابهی برای خوشه‌های دیگر انجام‌شده است که نتایج در **جدول ۳** گزارش‌شده است. ستون تحمل ریسک نشان‌دهنده ارزیابی ذهنی تحمل ریسک بر اساس نتایج هر یک از خوشه‌ها است. بنابراین با استفاده از خوشه‌بندی، ما کلیه افراد سرمایه‌گذار را به ۷ خوشه با خصوصیات رفتاری مختلف و ترجیحات ریسکی متفاوت دسته‌بندی نمودیم؛ بنابراین، زمانی که یک فرد جدید برای سرمایه‌گذاری به سرمایه‌گذار مراجعه می‌نماید، می‌تواند بررسی کند که در کدام‌یک از این طبقات قرار می‌گیرد و به آن فرد توصیه سرمایه‌گذاری‌ای، متناسب با تیپ شخصیتی فرد و میزان پذیرش ریسک آن فرد ارائه نماید.

جدول ۳- طبقه‌بندی افراد بر اساس ارزیابی ذهنی و نتایج خوشه‌بندی.

Table 3- Classification of individuals based on subjective evaluation and clustering results.

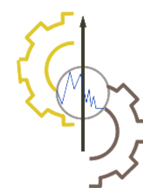
Cluster	Features	Risk Capacity
Cluster 0	Low Age, High Nelworth and Income, Less risky life category, willingness to spend more	High
Cluster 1	High Age, low net worth and Income, highly risky life category, willing ness to spend more, low education	High
Cluster 2	High Age, low net worth and Income, highly risky life category, willing ness to spend more, low education, own house	Medium
Cluster 3	Low Age, very low income and net worth, high willingness to take risky, many kids	Low
Cluster 4	Medium Age, very high income and net worth, high willingness to take risky, many kids, own house	High
Cluster 5	Low Age, very low income and net worth, high willingness to take risky, no kids	Medium
Cluster 6	Low Age, medium income and net worth, high willingness to take risky, many kids, own house	Low

بدین منظور چارچوب بومی در **شکل ۹** برای ایران پیشنهاد می‌شود:



شکل ۹- مدل بومی‌سازی پیشنهادی برای ایران.

Figure 9- Proposed localization model for Iran.



نوراحمدی و صادقی / مدیریت نوآوری و راهبردهای عملیاتی، دوره ۲، شماره ۲، تابستان ۱۴۰۰، صفحه: ۱۹۴-۱۸۰

طبق چارچوب پیشنهادی در **شکل ۹**، ابتدا فرد سرمایه‌گذار پرسشنامه تکمیل می‌کند (برای این انجام این کار می‌توانند از پرسشنامه ریسک سازمان بورس که برای همین منظور تعبیه‌شده است استفاده نمایند)، سپس با استفاده از خروجی پرسشنامه بررسی شود که سرمایه‌گذار در کدام دسته قرار می‌گیرد و استراتژی متناسب با تیپ شخصیتی خودش به او ارائه شود. گام بعدی این پژوهش نحوه تخصیص سرمایه‌گذاران به طبقات مختلف دارایی است که می‌توان از روش‌های مختلف بهینه‌سازی برای تعیین میزان وزن سرمایه‌گذاری در هر خوشه استفاده نمود.

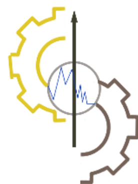
۵- نتیجه‌گیری

شناخت مشتریان و ارائه پیشنهادها سرمایه‌گذاری شخصی به آن‌ها، یک جنبه ضروری از یک استراتژی مشورتی سودمند و مؤثر است. بسیاری از خانوارها برای راهنمایی سرمایه‌گذاری به مشاوران مالی اعتماد می‌کنند. تجزیه‌وتحلیل داده‌های هوشمند یکی از زمینه‌های هوش مصنوعی است که مسئله یادگیری سیستم‌های اتوماتیک را بدون برنامه صریح حل می‌نماید. سیستم‌های توصیه‌کننده ابزارهای نرم افزاری هستند که برای ارائه پیشنهادها به کاربران بر اساس نیازهای آن‌ها استفاده می‌شوند. اگرچه هنوز این سیستم‌ها نسبتاً جدید هستند ولی در دهه اخیر توسعه بسیار خوبی نموده‌اند. یکی از عواملی که ارتباط زیادی با سیستم‌های توصیه‌کننده مرتبط است این است که بتوانیم توصیه-هایی متناسب با تیپ شخصیتی افراد به آن ارائه بدهیم. درعین‌حال، یکی از مسائلی که همواره ذهن مدیران سرمایه‌گذاری و توصیه گران سهام به افراد مختلف را درگیر می‌کند، مسئله تشکیل پرتفوی سهام با توجه به میزان و تمایل پذیرش ریسک سرمایه‌گذاران است. بدین منظور مدیران سرمایه‌گذاری باید بتوانند افراد مختلف سرمایه‌گذاران را با توجه به خصوصیات ریسک‌پذیری و ویژگی رفتاری و تمایل به ریسک به خوشه‌هایی تقسیم‌بندی نموده و سپس برای هر خوشه سبد سهام متناسب با همان ریسک را پیشنهاد نمایند. خوشه‌بندی سری‌های زمانی یکی از مؤثرترین روش‌ها در ارائه اطلاعات در حوزه‌های گوناگون است. خوشه‌بندی یکی از تکنیک‌های داده‌کاوی و شاخه‌ای از یادگیری ماشین بدون نظارت است که بدون هیچ‌دانشی در رابطه با تعداد و ویژگی گروه‌ها، بر اساس یک معیار شباهت، داده‌ها را گروه‌بندی می‌کند. در این پژوهش با استفاده از روش‌های خوشه‌بندی به خوشه‌بندی افراد سرمایه‌گذار می‌پردازیم. بدین منظور از دو روش خوشه‌بندی *K-means* و روش انتشار وابستگی استفاده نمودیم و با توجه به نتیجه ارزیابی ضریب *silhouette*، روش *k-means* برای خوشه‌بندی نهایی انتخاب شد. با توجه به اینکه در ایران ما محدودیت دیتا داریم و امکان دسترسی به اطلاعات مشتری‌های شرکت‌های سبد گردانی وجود ندارد، بنابراین برای این پژوهش از داده‌های مربوط به پرسشنامه تأمین مالی مصرف‌کننده *(SCF)* است که توسط شورای فدرال رزو انجام‌شده است[۱]، استفاده نمودیم. یکی از کلیدهای اصلی این پژوهش رویکرد درک شهودی خوشه‌ای است. ما برای درک رفتار مورد انتظار از یک عضو خوشه با تفسیر کیفی مقادیر متوسط متغیرها در هر خوشه از تکنیک‌های مصورسازی استفاده کردیم. کارایی روش خوشه‌بندی را در کشف گروه‌های بصری طبیعی سرمایه‌گذاران مختلف بر اساس تحمل ریسک آن‌ها نشان دادیم. با توجه به اینکه الگوریتم‌های خوشه‌بندی

---

۱ https://www.federalreserve.gov/econres/scf_2009p.htm

می‌توانند سرمایه‌گذاران را بر اساس عوامل مختلفی مانند سن، درآمد و تحمل ریسک با موفقیت گروه‌بندی کنند، می‌تواند به مدیران سبد سهام برای درک رفتار سرمایه‌گذار و استانداردسازی تخصیص پرتفوی و تعادل مجدد بین خوشه‌ها و سرمایه‌گذاری مورداستفاده قرار گیرد.



تعارض منافع

تعارض منافعی بین نویسندگان در نگارش این پژوهش وجود ندارد.

منابع

Aghabozorgi, S., Shirkhorshidi, A. S., & Wah, T. Y. (2015). Time-series clustering–a decade review. *Information systems*, *53*, 16-38. https://doi.org/10.1016/j.is.2015.04.007

Chandra, B., Gupta, M., & Gupta, M. P. (2008, October). A multivariate time series clustering approach for crime trends prediction. *2008 IEEE international conference on systems, man and cybernetics* (pp. 892-896). IEEE. DOI: 10.1109/ICSMC.2008.4811393

Chiu, B., Keogh, E., & Lonardi, S. (2003, August). Probabilistic discovery of time series motifs. *Proceedings of the ninth ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 493-498). https://doi.org/10.1145/956750.956808

Dose, C., & Cincotti, S. (2005). Clustering of financial time series with application to index and enhanced index tracking portfolio. *Physica A: statistical mechanics and its applications*, *355*(1), 145-151. https://doi.org/10.1016/j.physa.2005.02.078

Durante, F., Pappadà, R., & Torelli, N. (2014). Clustering of financial time series in risky scenarios. *Advances in data analysis and classification*, *8*(4), 359-376. https://doi.org/10.1007/s11634-013-0160-4

Ezugwu, A. E., Shukla, A. K., Agbaje, M. B., Oyelade, O. N., Jose-Garcia, A., & Agushaka, J. O. (2021). Automatic clustering algorithms: a systematic review and bibliometric analysis of relevant literature. *Neural computing and applications*, *33*(11), 6247-6306. https://doi.org/10.1007/s00521-020-05395-4

Fu, T. C., Chung, F. L., Ng, V., & Luk, R. (2001, August). Pattern discovery from stock time series using self-organizing maps. *Workshop notes of KDD2001 workshop on temporal data mining* (Vol. 1). https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.124.3646&rep=rep1&type=pdf

Graves, D., & Pedrycz, W. (2010). Proximity fuzzy clustering and its application to time series clustering and prediction. *10th international conference on intelligent systems design and applications* (pp. 49-54). IEEE. DOI: 10.1109/ISDA.2010.5687293

He, W., Feng, G., Wu, Q., He, T., Wan, S., & Chou, J. (2012). A new method for abrupt dynamic change detection of correlated time series. *International journal of climatology*, *32*(10), 1604-1614. https://doi.org/10.1002/joc.2367

Jain, P., & Jain, S. (2019). Can machine learning-based portfolios outperform traditional risk-based portfolios? The need to account for covariance misspecification. *Risks*, *7*(3), 74. https://doi.org/10.3390/risks7030074

Keogh, E., Lonardi, S., & Chiu, B. Y. C. (2002, July). Finding surprising patterns in a time series database in linear time and space. *Proceedings of the eighth ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 550-556). https://doi.org/10.1145/775047.775128

Leng, M., Lai, X., Tan, G., & Xu, X. (2009, August). Time series representation for anomaly detection. *2nd IEEE international conference on computer science and information technology* (pp. 628-632). IEEE. DOI: 10.1109/ICCSIT.2009.5234775

León, D., Aragón, A., Sandoval, J., Hernández, G., Arévalo, A., & Niño, J. (2017). Clustering algorithms for risk-adjusted portfolio construction. *Procedia computer science, ICCS*, *108*, 1334-1343. https://doi.org/10.1016/j.procs.2017.05.185

Lohre, H., Rother, C., & Schäfer, K. A. (2020). Hierarchical risk parity: accounting for tail dependencies in multi-asset multi-factor allocations. *Machine learning for asset management: new developments and financial applications*, 329-368. https://doi.org/10.1002/9781119751182.ch9

MacQueen, J. (1967, June). Some methods for classification and analysis of multivariate observations. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* (Vol. 1, No. 14, pp. 281-297). University of California.

Musto, C., Semeraro, G., Lops, P., De Gemmis, M., & Lekkas, G. (2015). Personalized finance advisory through case-based recommender systems and diversification strategies. *Decision support systems*, *77*, 100-111.

Polz, P. M., Hortnagl, E., & Prem, E. (2003). *Processing and clustering time series of mobile robot sensory data (OFAI-TR-2003-10)*. Austrian research institute for artificial intelligence: systemic intelligence for growingup artefacts that live-SIGNAL.

Raffinot, T. (2017). Hierarchical clustering-based asset allocation. *The journal of portfolio management*, *44*(2), 89-99. DOI: https://doi.org/10.3905/jpm.2018.44.2.089

Raffinot, T. (2018). The hierarchical equal risk contribution portfolio. *Available at SSRN 3237540*.

Rai, P., & Singh, S. (2010). A survey of clustering techniques. *International journal of computer applications*, *7*(12), 1-5.

Ren, F., Lu, Y. N., Li, S. P., Jiang, X. F., Zhong, L. X., & Qiu, T. (2017). Dynamic portfolio strategy using clustering approach. *PloS one*, *12*(1), e0169299. https://doi.org/10.1371/journal.pone.0169299

Saxena, A., Prasad, M., Gupta, A., Bharill, N., Patel, O. P., Tiwari, A., ... & Lin, C. T. (2017). A review of clustering techniques and developments. *Neurocomputing*, *267*, 664-681. https://doi.org/10.1016/j.neucom.2017.06.053

Sfetsos, A., & Siriopoulos, C. (2004). Time series forecasting with a hybrid clustering scheme and pattern recognition. *IEEE transactions on systems, man, and cybernetics-part A: systems and humans*, *34*(3), 399-405. DOI: 10.1109/TSMCA.2003.822270

Statman, M. (2018). Behavioral finance lessons for asset managers. *The journal of portfolio management*, *44*(7), 135-147.

Tatsat, H., Puri, S., & Lookabaugh, B. (2020). *Machine learning and data science blueprints for finance: from building trading strategies to robo-advisors using python.* O'Reilly media.

Thuraisingham, B. M., & Ceruti, M. G. (2000, October). Understanding data mining and applying it to command, control, communications and intelligence environments. *Proceedings 24th annual international computer software and applications conference. COMPSAC2000* (pp. 171-175). IEEE. DOI: 10.1109/CMPSAC.2000.884710

Tola, V., Lillo, F., Gallegati, M., & Mantegna, R. N. (2008). Cluster analysis for portfolio optimization. *Journal of economic dynamics and control*, *32*(1), 235-258. https://doi.org/10.1016/j.jedc.2007.01.034

Wang, H., Wang, W., Yang, J., & Yu, P. S. (2002, June). Clustering by pattern similarity in large data sets. *Proceedings of the 2002 ACM SIGMOD international conference on management of data* (pp. 394-405). https://doi.org/10.1145/564691.564737

Williams, B., Onsman, A., & Brown, T. (2010). Exploratory factor analysis: a five-step guide for novices. *Australasian journal of paramedicine*, *8*(3), 1-13.

